

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-227917

(43)Date of publication of application : 15.08.2000

(51)Int.Cl. G06F 17/30
G06F 3/00

(21)Application number : 11-028101

(71)Applicant : AGENCY OF IND
SCIENCE &
TECHNOL

(22)Date of filing : 05.02.1999 (72)Inventor : AIZONO TOSHIKO

KAJI HIROYUKI

MORIMOTO

YASUTSUGU

YAMAZAKI

NORIYUKI

IIDA KEIKO

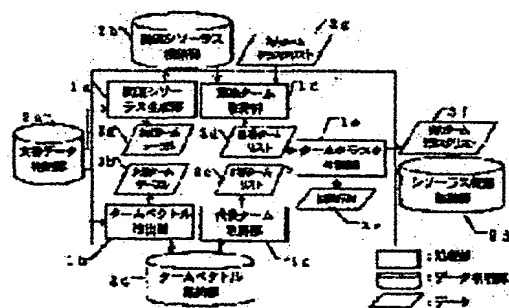
UCHIDA YASUHIKO

(54) THESAURUS BROWSING SYSTEM AND METHOD THEREFOR
AND RECORDING MEDIUM RECORDING ITS PROCESSING PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To improve the use efficiency of a related thesaurus by permitting a user to select any set, extracting a term related with a general term belonging to the respective representative terms of the selected set from the thesaurus and generating and displaying several sets from the extracted terms.

SOLUTION: The set of terms in document data stored in a document data storage part 2a is stored in a cooccurrence term table 3a. The term extracted from a document and the frequency are stored in the document term table 3b for every document. A related thesaurus is generated from document data in a related thesaurus generation part 1a. The term vectors of the respective documents are extracted from a term vector extraction part 1b and the view of the related thesaurus is



generated. A term cluster list is selected from it and the related term of the term belonging to the list is extracted. Several sets are generated from the extracted terms and they are displayed.

LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's
decision of rejection]

[Kind of final disposal of application
other than the examiner's decision
of rejection or application
converted registration]

[Date of final disposal for
application]

[Patent number]

[Date of registration]

[Number of appeal against
examiner's decision of rejection]

[Date of requesting appeal against
examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-227917
(P2000-227917A)

(43) 公開日 平成12年8月15日 (2000.8.15)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/30		G 0 6 F 15/403	3 2 0 D 5 B 0 7 5
3/00	6 5 6	3/00	6 5 6 C 5 E 5 0 1
		15/401	3 1 0 D
		15/403	3 8 0 E

審査請求 未請求 請求項の数15 O L (全 20 頁)

(21) 出願番号 特願平11-28101

(22) 出願日 平成11年2月5日 (1999.2.5)

(71) 出願人 000001144
工業技術院長
東京都千代田区霞が関1丁目3番1号
(72) 発明者 相蘭 敏子
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内
(72) 発明者 梶 博行
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内
(74) 代理人 100077274
弁理士 磯村 雅俊

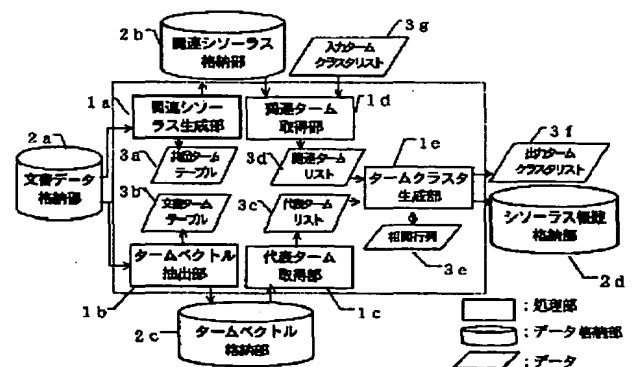
最終頁に続く

(54) 【発明の名称】 シソーラスブラウジングシステムと方法およびその処理プログラムを記録した記録媒体

(57) 【要約】

【課題】 従来の技術はネットワーク型の関連シソーラスの一部のみを表示するものであり、シソーラスの全体構造を表示提供することができない。

【解決手段】 一般的なタームから構成されるシソーラスの概観から、より限定的なタームを含む部分構造へユーザをナビゲートするシソーラスブラウジング技術を提供し、ユーザは、関連シソーラスの概観を容易に把握でき、更に興味を持ったタームクラスタのズームングにより、効率的に関連シソーラスをブラウジングすることができる。



【特許請求の範囲】

【請求項 1】 関連関係にある複数のタームを格納したシソーラスの検索・表示を行なうシソーラスブラウジングシステムであって、上記シソーラスの概観を表す複数の一般的タームを抽出してクラスタリングした概観タームクラスタを表示するシソーラス概観生成手段と、ユーザが選択した上記概観タームクラスタに属する各一般的タームに関連関係のあるタームを抽出してクラスタリングした限定タームクラスタを表示するシソーラスズームリング手段とを有することを特徴とするシソーラスブラウジングシステム。

【請求項 2】 請求項 1 に記載のシソーラスブラウジングシステムにおいて、ユーザが選択した上記限定タームクラスタに属する各限定タームに関連関係のあるタームを抽出してクラスタリングし、新たに得られる限定タームクラスタを表示する第 2 のシソーラスズームリング手段を有することを特徴とするシソーラスブラウジングシステム。

【請求項 3】 請求項 1、もしくは、請求項 2 のいずれかに記載のシソーラスブラウジングシステムにおいて、上記シソーラス概観生成手段は、関連関係にあるタームから上記一般的タームとなる代表タームを抽出する代表ターム取得手段と、該代表ターム取得手段で抽出した代表タームから関連度の高いものをまとめて上記概観タームクラスタを生成するタームクラスタ生成手段とを有することを特徴とするシソーラスブラウジングシステム。

【請求項 4】 請求項 1 から請求項 3 のいずれかに記載のシソーラスブラウジングシステムにおいて、上記シソーラスズームリング手段は、上記ユーザが選択した概観タームクラスタに属する一般的タームと関連関係にあるタームを上記シソーラスから抽出する関連ターム取得手段と、該関連ターム取得手段で抽出したターム間の関連度を上記シソーラスから取得し、関連度の高いタームをまとめて上記限定タームクラスタを生成するタームクラスタ生成手段とを有することを特徴とするシソーラスブラウジングシステム。

【請求項 5】 関連関係にある複数のタームを格納したシソーラスの検索・表示を行なうシソーラスブラウジングシステムであって、ユーザが入力したタームに関連関係のあるタームを上記シソーラスから抽出する関連ターム取得手段と、該関連ターム取得手段で抽出したターム間の関連度を上記シソーラスから取得し、関連度の高いタームをまとめてタームクラスタを生成するタームクラスタ生成手段とを有し、上記タームクラスタ生成手段で生成したタームクラスタを表示することを特徴とするシソーラスブラウジングシステム。

【請求項 6】 請求項 5 に記載のシソーラスブラウジングシステムにおいて、ユーザが選択したタームクラスタに属する各タームに関連関係のあるタームを抽出してクラスタリングした限定タームクラスタを表示するシソー

ラスズームリング手段を有することを特徴とするシソーラスブラウジングシステム。

【請求項 7】 関連関係にある複数のタームを格納したシソーラスの検索・表示を行なう装置のシソーラスブラウジング方法であって、上記シソーラスの概観を表す複数の一般的タームを抽出してクラスタリングした概観タームクラスタを表示するシソーラス概観生成処理ステップと、ユーザが選択した上記概観タームクラスタに属する各一般的タームに関連関係のあるタームを抽出してクラスタリングした限定タームクラスタを表示するシソーラスズームリング処理ステップとを有することを特徴とするシソーラスブラウジング方法。

【請求項 8】 請求項 7 に記載のシソーラスブラウジング方法において、ユーザが選択した上記限定タームクラスタに属する各限定タームに関連関係のあるタームを抽出してクラスタリングし、新たに得られる限定タームクラスタを表示する第 2 のシソーラスズームリング処理ステップを有することを特徴とするシソーラスブラウジング方法。

【請求項 9】 請求項 7、もしくは、請求項 8 のいずれかに記載のシソーラスブラウジング方法において、上記シソーラス概観生成処理ステップは、関連関係にあるタームから上記一般的タームとなる代表タームを抽出する代表ターム取得処理ステップと、該代表ターム取得処理ステップで抽出した代表タームから関連度の高いものをまとめて上記概観タームクラスタを生成するタームクラスタ生成処理ステップとを有することを特徴とするシソーラスブラウジング方法。

【請求項 10】 請求項 7 から請求項 9 のいずれかに記載のシソーラスブラウジング方法において、上記シソーラスズームリング処理ステップは、上記ユーザが選択した概観タームクラスタに属する一般的タームと関連関係にあるタームを上記シソーラスから抽出する関連ターム取得処理ステップと、該関連ターム取得処理ステップで抽出したターム間の関連度を上記シソーラスから取得し、関連度の高いタームをまとめて上記限定タームクラスタを生成するタームクラスタ生成処理ステップとを有することを特徴とするシソーラスブラウジング方法。

【請求項 11】 請求項 7 から請求項 10 のいずれかに記載のシソーラスブラウジング方法において、上記概観タームクラスタに属するタームの追加・削除を、ユーザからの編集指示に基づき行なうタームクラスタ編集処理ステップを有し、ユーザが編集して選択した上記概観タームクラスタに対して上記シソーラスズームリング処理ステップでの処理を行なうことを特徴とするシソーラスブラウジング方法。

【請求項 12】 請求項 7 から請求項 11 のいずれかに記載のシソーラスブラウジング方法において、上記シソーラスズームリング処理ステップで表示した限定タームクラスタに属するタームの内、ユーザが選択した上記概観ター

ームクラスに属していたタームを識別表示するタームクラス表示処理ステップを有することを特徴とするシソーラスブラウジング方法。

【請求項13】 関連関係にある複数のタームを格納したシソーラスの検索・表示を行なう装置のシソーラスブラウジング方法であって、ユーザが入力したタームに関連関係のあるタームを上記シソーラスから抽出する関連ターム取得処理ステップと、該関連ターム取得手段で抽出したターム間の関連度を上記シソーラスから取得し、関連度の高いタームをまとめてタームクラスタを生成するタームクラスタ生成処理ステップとを有し、上記タームクラスタ生成手段で生成したタームクラスタを表示することを特徴とするシソーラスブラウジング方法。

【請求項14】 請求項13に記載のシソーラスブラウジング方法において、ユーザが選択したタームクラスタに属する各タームに関連関係のあるタームを抽出してクラスタリングした限定タームクラスタを表示するシソーラスズーム処理ステップを有することを特徴とするシソーラスブラウジング方法。

【請求項15】 コンピュータが読み取り可能なプログラムおよびデータを記録する記録媒体であって、請求項7から請求項14のいずれかに記載の各ステップの処理をコンピュータに実行させるためのプログラムを記録したことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、シソーラスに格納された各タームの表示技術に係わり、特に、ネットワーク型のシソーラスに格納されたタームの検索の効率化を図るのに好適なシソーラスブラウジングシステムと方法およびその処理プログラムを記録した記録媒体に関するものである。

【0002】

【従来の技術】以下、まず、このシソーラスの概要についてJIS規格を引用して簡単に述べ、次に、本発明が対象とするターム（関連ターム）を格納したシソーラスの特徴について説明し、その後、当該シソーラスをブラウジングするための従来の技術を示す。

【0003】JIS X 0901「シソーラスの構成及びその作成方法」によれば、シソーラスとは「統制された索引言語の語彙であって、あらかじめ概念間の先験的な関係を明示するように組織化したもの」である。ここで「索引言語の語彙」、すなわち「索引語」とは、「ある概念を名詞又は名詞相当の句の形で適切に表現したもの」（JIS X 0901）であり、一般には「ターム」と呼ばれることが多い。よって、以下「ターム」と呼ぶ。

【0004】シソーラスに記述されるターム間の関係には、「同義関係」、「階層関係」及び「関連関係」がある（JIS X 0901）。「同義関係」とは、同じ概

念を指すターム間の関係である。例えば、「カメラ」と「写真機」などである。「階層関係」とは、上位概念と下位概念間の関係であり、種類関係（クラスとそのメンバの関係、例：「霊長類」と「サル」）、全体部分関係（部分の名称とそれが属する全体の名称、例：「消化器」と「腸」）、及び例示関係（カテゴリとその事例の関係、例：「高速鉄道」と「新幹線」）の3種類がある。このような階層関係にあるタームを格納したシソーラスを以下「階層シソーラス」と呼ぶ。

【0005】これに対して「関連関係」は、「同義関係及び階層関係以外の関連を持つ語の関係」（JIS X 0901）であり、意味が部分的に重複している場合（例：「乗用車」と「自家用車」）や、一方のタームが他方のタームを強く含意する場合（例：「出版」と「図書」）がある。このような関連関係にあるターム、すなわち関連タームを格納したシソーラス（以下、「関連シソーラス」）の特徴として、ネットワーク型であるという点が挙げられる。例を用いて説明する。

【0006】図2は、関連シソーラスの一例を示す説明図である。図2中、ノードはターム、ノード間のリンクは関連関係を示す。本図2に示すように関連シソーラスでは、「公定歩合」と「利率」に関連関係があり、さらに「利率」と「ローン」が関連関係で結ばれ、更に「ローン」と「貸し渋り」に関連関係がある。このように関連シソーラスでは、タームが互いに関連を持ってネットワークを形成している。

【0007】このような関連シソーラスをブラウジングする従来の技術として、次の2つがある。まず、特開平5-233717号公報に記載の「情報検索装置」は、ユーザがタームを入力すると、ユーザが指定した範囲内の関連度を持つ関連タームを検索して表示する。関連度の範囲は、関連度の強さによってN段階ある（例：「大中小の3段階」）。これにより、ユーザは関連度の低い関連タームも効率よく参照することが出来る。

【0008】また、特開平9-44525号公報に記載の「データ検索装置」は、ユーザがタームを入力すると、入力されたタームとの関連度の強さに従って関連タームの表示位置を決定する。より具体的には、関連シソーラスを図2に示すようなネットワーク構造で2次元的に表示する方法に関し、ユーザが入力したタームとの関連度が強いタームほど入力タームの近くに表示する。これにより、ユーザは入力したタームとの関連度の強さをビジュアルに把握できる。

【0009】これらの従来技術は、ユーザが関連シソーラスの一部分のみに興味を持つとき有用である。しかし、その一方で、（1）関連シソーラスの概観をユーザに提供することができない、（2）関連シソーラスの一部分に興味を持っている場合でも、所望のタームに辿り着くまでに手間がかかる、という問題点がある。以下、それぞれについて説明する。

【0010】(1) 関連シソーラスの概観をユーザに提供することができない：上述した従来の技術は、ユーザが入力したキータームの関連タームのみを表示する。その一方で、関連シソーラスの全体の概観を掴みたいというユーザの要求がある。例えば、関連シソーラスを参照したいものの、特に何について参照したいかは実際に関連シソーラスを調べなければつきりと分からないときがある。

【0011】このような場合ユーザは、まずどのようなタームが格納されているか関連シソーラスの概観を掴んでから興味のある方向へネットワークの部分構造を辿っていきたいと思うであろう。例えば、階層シソーラスでは、ユーザはその最上位のタームのリストを参照すれば概観を掴むことができる。これに対して、上記従来技術では、予め関連シソーラスのネットワークのどの部分を参照するかユーザが決めておく必要がある。そのため、関連シソーラスの概観を掴んでからシソーラスをブラウジングしたいというユーザの要求に応えることができない。

【0012】(2) 所望のタームに辿り着くまでに手間がかかる：上述のように従来の技術では、ユーザによるキータームの入力が必要である。その一方で、適切なキーの入力は、簡単ではないことが知られている。特に検索対象の関連シソーラスに熟知していないユーザにとって、入力するタームとは適当に思い付いたタームであることが多い。例えば「経済」について知りたいとき、「経済」をキータームとして入力してしまう、などである。

【0013】このように思い付いたタームをキーとして入力したとき、ユーザの興味に一致する関連タームが直接得られる可能性は低い。そのため表示された関連タームの中からキーとなるタームを選択して、さらに関連タームを検索し辿っていくことが必要であろう。上述のように関連シソーラスはネットワーク型のシソーラスなので、関連タームを次々と辿っていくうちにユーザが“迷子”、すなわちどのようにシソーラスを辿ったか分からなくなることもある。このように、思い付いたタームを適当に入れてシソーラスを調べるのは効率が悪い。

【0014】

【発明が解決しようとする課題】解決しようとする問題は、従来の技術では、関連シソーラスの全体構造（概観）をユーザに提供することができない点と、ユーザが関連シソーラスの一部分に興味を持っている場合でも所望のタームに辿り着くまでに手間がかかってしまう点である。

【0015】本発明の目的は、これら従来技術の課題を解決し、関連シソーラスの利用効率を向上させることを可能とするシソーラスブラウジングシステムと方法およびその処理プログラムを記録した記録媒体を提供することである。

【0016】

【課題を解決するための手段】上記目的を達成するため、本発明のシソーラスブラウジングシステムと方法は、従来技術のようにネットワーク型の関連シソーラスの一部分のみを表示するものではなく、関連シソーラスに格納されたタームの内、この関連シソーラスの全体構造（概観）を示すようなターム（代表ターム）の集合を幾つか生成して表示する。そして、ユーザがいずれかの集合を選択した場合、選択された集合の各代表タームに属する一般的なタームに関連するタームをシソーラスから抽出し、さらに、抽出した各タームから同様に、幾つかの集合を生成して表示する。

【0017】これにより、関連シソーラス中、一般的なタームで構成された全体構造から、より限定的なタームも含む部分構造へとユーザをナビゲートすることができる。例えば、ナビゲーションの初期状態としてシソーラスの概観を示すタームの集合を表示した画面から、ユーザは興味を持った部分（タームの集合）を選びズームインを指示することにより、興味を持った部分について、より限定的なタームを含む詳細構造を参照することができる。

【0018】

【発明の実施の形態】以下、本発明の実施の形態を、図面により詳細に説明する。図1は、本発明のシソーラスブラウジングシステムの本発明に係る構成の第1の実施形態例を示すブロック図であり、図2は、図1におけるシソーラスブラウジングシステムの処理対象の関連シソーラスの内容を示す説明図であり、図3は、図1におけるシソーラスブラウジングシステムの本発明に係る処理の概要を示す説明図であり、図4は、図1におけるシソーラスブラウジングシステムのハードウェア構成例を示すブロック図である。

【0019】まず、図2および図3を用いて、本例のシソーラスブラウジングシステムの概要を説明する。本例のシソーラスブラウジングシステムは、特に経済分野に属する文書のデータベースから自動的に図2に示す関連シソーラスを生成し、その概観を、図3における一般的なタームとしてまとめられた複数のタームの集合（タームクラスタ）単位で表示してユーザに提供する。

【0020】この図3により、経済分野における関連シソーラスのブラウジングを説明する。図3中、最上位は、一般的なタームの集合から構成されるシソーラスの概観である。図3に示すようにシソーラスの概観は、「税制」、「所得」、「国税庁」等を含むタームの集合、「日銀」、「外貨」、「円」、「為替」を含むタームの集合等から構成される。

【0021】これらのうち、ユーザが「日銀」、「外貨」、「円」、「為替」を含むタームの集合を選択すると、選択したタームの集合に含まれるタームの関連タームで、より限定的なタームを含む集合が表示される。そ

の一例として図3の中段には「円」、「円売り」、「円買い」、「東京外為市場」を含むタームの集合、「外貨」、「ドル」、「IMF」、「ユーロ」を含むタームの集合等が表示されている。

【0022】これらのうちからユーザが更に興味を持ったタームの集合を選択すれば、より限定的なタームから成るタームの集合を参照することができる。すなわち、図3中下段の「円買い」、「円高」、「貿易黒字」、「不均衡」を含むタームの集合等がその一例である。

【0023】このように一般的なタームから構成されるタームの集合を参照することにより、ユーザは関連シソーラスの概観を掴むことが可能となる。更に、ユーザが選択した一般的なタームの集合からより限定的なタームを含む部分集合を表示するインタラクティブな処理（以下「ズームング」と呼ぶ。）により、効率良くシソーラスをブラウジングすることができる。

【0024】このようなシソーラスのブラウジングを実現するため、本例では、シソーラス概観生成機能とタームクラスタのズームング機能を設ける。シソーラス概観生成機能は、関連シソーラスの概観を生成するものであり、次の2つの処理機能からなる。すなわち、(a)シソーラスに格納されている関連関係が対応している分野から一般的なタームを抽出する代表ターム抽出処理機能と、(b)この代表ターム抽出処理機能で抽出したタームのうち関連度の高いものをまとめてクラスタを生成するタームクラスタ生成処理機能である。

【0025】以下、シソーラスが対応する分野における一般的なタームで、シソーラス概観の要素となるタームを「代表ターム」、タームの集合からクラスタを生成する処理を「タームクラスタリング」、タームを構成要素とするクラスタを「タームクラスタ」と呼ぶ。

【0026】タームクラスタのズームング処理機能は、一般的なタームの集合からより限定的なタームを含む集合を生成して表示するものであり、次の2つの処理機能からなる。すなわち、(a)ユーザがタームクラスタを選択すると当該タームクラスタに属するタームの関連タームを取得する関連ターム取得処理機能と、(b)この関連ターム取得処理機能で取得した関連タームをクラスタリングするタームクラスタ生成処理機能である。

【0027】このようなシソーラスのブラウジングを実現するため各機能を実装するためのハードウェア構成を図4を用いて説明する。本図4に示すように、本例のシソーラスブラウジングシステムは、CPU (Central Processing Unit) 1、ハードディスク2、メモリ3、ディスプレイ4a、ディスプレイ制御部4b、キーボード5a、キーボード制御部5b、マウス6a、マウス制御部6b、及びバス7から構成される。

【0028】CPU1は、プログラムに基づき、データの入出力、読み込み、格納及び各種処理を実行して、本発明に係わるシソーラスブラウジング処理を行なう。ハ

ードディスク2は、データを保存する装置、メモリ3は、プログラム及びデータをロードして記憶する装置である。ディスプレイ4aは、ユーザにデータを表示する装置であり、ディスプレイ制御部4bによって制御される。キーボード5a及びマウス6aはユーザからの入力を受け付ける装置であり、それぞれキーボード制御部5b及びマウス制御部6bによって制御される。バス7は、各構成要素間にデータを受け渡す。

【0029】このようなハードウェア構成において、図示していない光ディスク等の記録媒体から本発明のシソーラスブラウジング方法に係わる処理プログラムをハードディスク2に格納し、メモリ3にロードして起動することにより、図1に示すシソーラスブラウジングシステムが構成される。

【0030】図1におけるシソーラスブラウジングシステムは、次の3つのモジュールに分類される。すなわち、各種処理を実行する処理部（図中、矩形で示す）、データを格納するデータ格納部（図中、ドラム形で示す）、及びデータ（図中、平行四辺形で示す）である。以下、それぞれについて説明する。

【0031】まず処理部として、関連シソーラス生成部1a、タームベクトル抽出部1b、代表ターム取得部1c、関連ターム取得部1d、及びタームクラスタ生成部1eの5つがある。関連シソーラス生成部1aは、文書データ格納部2aに格納された文書データから関連シソーラスを生成する。

【0032】タームベクトル抽出部1bは、文書データ格納部2aに格納された文書データからタームベクトルを抽出する。代表ターム取得部1cは、タームベクトル格納部2cに格納されたタームベクトルから代表タームを取得する。関連ターム取得部1dは、ユーザが選択したタームクラスタに属するタームの関連タームを取得する。タームクラスタ生成部1eは、代表タームリスト3c又は関連タームリスト3dに格納されたタームをクラスタリングする。尚、これらの各部の詳細な処理手順についてはフローチャートを用いて後述する。

【0033】次にデータ格納部として、文書データ格納部2a、関連シソーラス格納部2b、タームベクトル格納部2c、及びシソーラス概観格納部2dの4つがある。文書データ格納部2aは、ある分野に属する文書のデータを格納する。例えば、経済面に掲載された新聞記事、或いはある分類番号を付与された特許の明細書などである。

【0034】関連シソーラス格納部2bは、文書データ格納部2aに格納された文書データから生成した関連シソーラスを格納する。タームベクトル格納部2cは、文書データ格納部2aに格納された文書データから抽出したタームベクトルを格納する。シソーラス概観格納部2dは、関連シソーラス格納部2bに格納された関連シソーラスの概観を格納する。

【0035】ここで各データ格納部の詳細な構成について図5～図8を用いて説明する。図5は、図1における文書データ格納部の構成例を示す説明図である。本図5に示すように文書データ格納部2aは、文書データ2a01から構成される。文書データ2a01には、文書のテキストデータを格納する。図5中第1行目には、経済に関する新聞記事の一例として「21日の東京外為市場の円相場は…」が格納されている。

【0036】図6は、図1における関連シソーラス格納部の構成例を示す説明図である。本図6に示すように関連シソーラス格納部2bは、タームX2b01、タームY2b02、及び関連度2b03から構成される。タームX2b01及びタームY2b02には、関連関係にあるタームを、関連度2b03にはその関連度を格納する。図6中、第1行目には「円」と「東京外為市場」が関連タームであり、その関連度は「11.5」であることが格納されている。尚、本例における関連度はタームの集合の相互情報量の値である。これについては後述する。

【0037】図7は、図1におけるタームベクトル格納部の構成例を示す説明図である。本図7に示すようにタームベクトル格納部2cは、文書ID2c01、及び重要タームリスト2c02から構成される。文書ID2c01には、文書データ格納部2aに格納された文書のIDを格納し、重要タームリスト2c02には当該文書に出現するタームのうち重要なもののリストを格納する。図7中、第1行目には文書ID1の文書の重要タームが「円高」、「東京外為市場」、「ドル安」等であることが格納されている。

【0038】図8は、図1におけるシソーラス概観格納部の構成例を示す説明図である。本図8に示すようにシソーラス概観格納部2dは、タームリスト2d01から構成される。タームリスト2d01には、タームクラスタに属するタームのリストを格納する。図8中、第1行目には、「景気」、「売り上げ」、「消費者」等から構成されるタームクラスタが格納されている。尚、図8に示すシソーラス概観は、一例として10個のタームクラスタからなるものとする。

【0039】以上、図5～図8を用いて図1に示すモジュール構成のうちデータ格納部につ次に、データについて説明する。図1に示すようにデータとして、共起タームテーブル3a、文書タームテーブル3b、代表タームリスト3c、関連タームリスト3d、相関行列3e、出力タームクラスタリスト3f、及び入力タームクラスタリスト3gがある。

【0040】共起タームテーブル3aは、文書データ格納部2aに格納された文書データ中、ある一定の範囲内で同時に出現するタームの組（タームの集合）を格納する。文書タームテーブル3bは、文書データ格納部2aに格納された文書から抽出したタームとその頻度を文書

ごとに格納する。代表タームリスト3cは、タームベクトル格納部2bから抽出した代表タームを格納する。関連タームリスト3dは、入力タームクラスタ3gに格納されたタームの関連タームを格納する。

【0041】相関行列3eは、代表タームリスト3c又は関連タームリスト3dに格納されたターム間の関連度を格納する。出力タームクラスタリスト3fは、代表タームリスト3c又は関連タームリスト3dに格納されたタームをクラスタリングした結果を格納する。入力タームクラスタリスト3gは、ユーザが選択したタームクラスタを格納する。

【0042】ここで、各データの詳細について図9～図15を用いて説明する。図9は、図1における共起タームテーブルの構成例を示す説明図である。本図9に示すように、共起タームテーブル3aは、タームX3a01、タームY3a02、及び共起頻度3a03から構成される。タームX3a01及びタームY3a02には、文書データ格納部2aに格納された文書データ中、ある一定の範囲内で同時に出現するタームの組（集合）を格納する。共起頻度3a03には、当該タームの組（集合）が同時に出現した頻度を格納する。

【0043】尚、「共起」とはあるタームが他のタームと一定の範囲内で同時に出現することであり、「共起頻度」とは同時に出現した回数を指す。更にあるタームと共起するタームを「共起ターム」と呼ぶ。ここでは一定の範囲の一例として「同じ文」を用いる。一例として図9中第1行目には、「円相場」と「東京外為市場」が259回、同じ文に出現したことが格納されている。

【0044】図10は、図1における文書タームテーブルの構成例を示す説明図である。本図10に示すように、文書タームテーブル3bは、文書ID3b01、ターム3b02、及び出現頻度3b03から構成される。文書ID3b01は、文書データ格納部2aに格納された文書のIDを格納する。ターム3b02には、当該IDの文書に出現したタームを、出現頻度3b03にはタームの当該文書内での出現頻度を格納する。図10中、第1行目には文書ID「1」の文書中、「東京外為市場」が2回出現したことが格納してある。

【0045】図11は、図1における代表タームリストの構成例を示す説明図である。図11(a)はタームベクトルから重要タームを取得しているときの状態を示し、図11(b)は重要タームの取得後ソートされた状態のリストを示す。図11(a)に示すように代表タームリスト3cは、代表ターム3c01及び文書数3c02から構成される。

【0046】代表ターム3c01には、タームベクトル格納部2bから抽出した代表タームを格納する。文書数3c02には、当該タームが重要タームとなった文書数を格納する。図11(a)中、第1行目には「為替」が451文書で重要タームであったことが格納されてる。

図11(a)のリストを文書数の降順にソートした状態が図11(b)であり、図11(b)中、第1行目には「株」が498文書で重要タームであったことが格納されている。

【0047】図12は、図1における関連タームリストの構成例を示す説明図である。図12(a)はユーザが選択したタームクラスタに属するタームの関連タームを取得しているときの状態を示し、図12(b)は関連タームの取得後ソートされた状態のリストを示す。図12(a)に示すように関連タームリスト3dは、関連ターム3d01及び順位3d02から構成される。

【0048】関連ターム3d01には、ユーザが選択したタームクラスタに属するタームの関連タームを格納する。順位3d02には、当該タームの関連タームのうち関連度が何番目に大きかったかを表す順位を格納する。図12(a)中第15番目には「東京外為市場」が順位1として格納されている。このことは、「東京外為市場」があるタームの関連タームのうち最も関連度が大きかったことを表す。

【0049】図12(a)のリストを順位の昇順でソートした状態が図12(b)である。この図12(b)中、第1行目から第10行目には入力されたタームクラスタに属する各タームと関連度が最も大きかったターム、すなわち順位1のタームが格納されている。

【0050】図13は、図1における相関行列の構成例を示す説明図である。本図13に示すように相関行列3eは、 $n \times n$ の行列である。ここで n は代表タームリスト3c又は関連タームリスト3dに格納されたタームの数である。図13に示す相関行列3e中、 i 行目 j 列目の要素は、代表タームリスト3c又は関連タームリスト3d中、 i 番目のタームと j 番目のタームの関連度である。

【0051】図13中、第1行目の第2列には関連度「7.3」が格納されている。尚、相関行列3eでは、行列の対角要素には同じ値が格納されている。すなわち、図13中、第2行目第1列には、第1行目第2列と同じ値「7.3」が格納されている。

【0052】図14は、図1における出力タームクラスタリストの構成例を示す説明図である。図14(a)は出力タームクラスタリスト3fの初期状態を示し、図14(b)はタームクラスタ生成後の状態を示す。図14(a)に示すように出力タームクラスタリスト3fは、最大関連度のクラスタID3f01、最大関連度3f02、及びタームリスト3f03から構成される。

【0053】出力タームクラスタリスト3fの各行は、1つのタームクラスタを表し、各行のインデックスがすなわちクラスタIDとなる。最大関連度のクラスタID3f01には、当該タームクラスタと最大の関連度をとるタームクラスタのIDを格納し、最大関連度3f02にはその最大関連度を格納する。タームクラスタ間の関

連度については後述する。タームリスト3f03には、当該タームクラスタに属するタームのリストを格納する。

【0054】図14(a)に示すように、出力タームクラスタリスト3fには、初期状態として1つのタームをタームリスト3f03に格納したタームクラスタが格納されている。これら1つのタームから成るタームクラスタを定められた数になるまでマージした状態が図14(b)である。この図14(b)中の第1行目には「IMF」、「ドル」、「外貨」、「ユーロ」などから構成されるタームクラスタが格納されている。

【0055】図15は、図1における入力タームクラスタリストの構成例を示す説明図である。図15に示すように入力タームクラスタリスト3gは、タームリスト3g01から構成される。出力タームクラスタリスト3fと同様入力タームクラスタリスト3gの各行は、1つのタームクラスタを表す。図15中の第1行目には「円」、「為替」、「日銀」、「外貨」などから構成されるタームクラスタが格納されている。

【0056】以上、本例のシソーラスブラウジングシステムのモジュール構成及びその構成要素の詳細について説明した。次に、本例のシソーラスブラウジングシステムの詳細な処理手順について図16～図26に示すフローチャートを用いて説明する。

【0057】まず、シソーラスブラウジングシステムの全体の処理手順について図16を用いて説明する。図16は、本発明のシソーラスブラウジング方法の処理手順例を示すフローチャートである。

【0058】本図16に示すように、本例のシソーラスブラウジング方法は、大きく次の2つの処理から構成される。すなわち、図16(a)に示すシソーラスブラウジング用データ生成処理と、図16(b)に示すシソーラスブラウジング処理である。前者はシソーラスブラウジングのためのデータを生成する処理であり、バッチ的に実行する。これに対して後者は、ユーザとインタラクティブに応答する処理であり、リアルタイムに実行する。以下、図16(a)及び図16(b)についてそれぞれ説明し、次に各ステップについて詳細なフローチャートを用いて説明する。

【0059】まず、図16(a)に示すシソーラスブラウジング用データ生成処理について説明する。図16

(a)に示すようにシソーラスブラウジング用データ生成処理では、まず文書データから関連シソーラスを生成し(ステップ101)、各文書のタームベクトルを抽出して(ステップ102)、関連シソーラスの概観を生成する(ステップ103)。

【0060】次に、図16(b)に示すシソーラスブラウジング処理について説明する。図16(b)に示すようにシソーラスブラウジング処理では、まずシソーラス概観格納部2dに格納されたシソーラス概観をユーザに

表示し（ステップ111）、次にユーザが表示されたタームクラスタリストを選択してズームングを指示すれば（ステップ112）、ユーザが選択したタームクラスタに属するタームの関連タームを取得する（ステップ113）。

【0061】そして、それらをクラスタリングし（ステップ114）、生成したタームクラスタをユーザに表示する（ステップ115）。ユーザからのシソーラスブラウジング終了の指示があれば（ステップ116）、処理を終了し、なければステップ112の処理に戻る。

【0062】以下、図16（a）におけるステップ101の関連シソーラス生成処理、ステップ102のタームベクトル抽出処理、ステップ103のシソーラス概観生成処理、及び図16（b）におけるステップ113の関連ターム取得処理について詳細に説明する。尚、ステップ114のタームクラスタ生成処理は、図16（a）に

相互情報量計算式

$$\text{rel}(T_i, T_j) = \log_2 \frac{P(T_i, T_j)}{P(T_i)P(T_j)} = \log_2 \frac{\frac{m}{M}}{\frac{n_i}{N} \cdot \frac{n_j}{N}} = \log_2 \frac{m \cdot N^2}{n_i \cdot n_j \cdot M}$$

$\text{rel}(T_i, T_j)$: ターム T_i とターム T_j の相互情報量度

n_i : ターム T_i の出現頻度

n_j : ターム T_j の出現頻度

N : ターム出現頻度総和

m : T_i, T_j の共起頻度

M : 共起頻度総和

【0065】図17は、図16（a）のシソーラスブラウジング用データ生成処理における関連シソーラス生成処理の詳細な処理手順例を示すフローチャートである。本図17に示すように関連シソーラス生成処理は、まず文書データ格納部2aに格納された文書の数を示す変数 i に1をセットして初期化し（ステップ1a01）、 i 番目の文書データを単語に分割しながら同じ文に出現したタームの組をすべて抽出し、共起タームテーブル3aに格納する（ステップ1a02）。

【0066】このとき、共起タームテーブル3aに同じタームの組があれば共起頻度を1増やし、同じタームの組がなければ共起頻度を1として格納する。次に i を1増やして（ステップ1a03）、 i が文書数より小さければ（ステップ1a04）ステップ1a02に戻る。

【0067】 i が文書数を超えれば、共起タームテーブル3aのインデックスを示す変数 j を1に初期化して（ステップ1a05）、共起タームテーブル3aの j 番目のタームの組の関連度を数1の式を用いて計算し、ある閾値以上の関連度を持つタームの組を関連シソーラス格納部2bに格納する（ステップ1a06）。次に j を1増やして（ステップ1a07）、 j が共起タームテーブル3aに格納されたタームの組数よりも小さければ（ステップ1a08）、ステップ1a06に戻り、 j が

示すシソーラス概観生成処理に含まれるタームクラスタ生成処理と同じなので、説明を省略する。

【0063】まず、関連シソーラス生成処理の詳細な手順の説明を行なう。文書データから共起タームを抽出して関連シソーラスを生成する処理は、特開平5-282367号公報（「関連キーワード自動生成装置」）、あるいは特開平8-161343号公報（「関連語辞書作成装置」）等で公知であるが、ここではその一例として、文書データから共起タームを抽出し、相互情報量に基づいて関連度を求める処理について説明する。

【0064】尚、相互情報量とは、「直接得られる事象 X （ここではターム）から推定される他の事象 Y の情報量」と定義されるが、ここではタームの組（タームの集合）の共起しやすさを正規化した値として考える。相互情報量を求める式を数1に示す。

【数1】

共起数を超えれば処理を終了する。

【0068】次に、タームベクトル抽出処理の詳細な手順を説明する。図18は、図16（a）のシソーラスブラウジング用データ生成処理におけるタームベクトル抽出処理の詳細な処理手順例を示すフローチャートである。本図18に示すようにタームベクトル抽出処理は、まず文書データ格納部2aに格納された文書の数を示す変数 i に1をセットして初期化し（ステップ1b01）、 i 番目の文書を単語に分割して各単語とその出現回数を文書タームテーブル3bに格納する（ステップ1b02）。

【0069】そして i を1増やして（ステップ1b03）、 i が文書数より小さければ（ステップ1b04）ステップ1b02に戻る。 i が文書数を超えれば、文書タームテーブル3bに格納されたタームそれぞれについて文書データベース全体での出現文書数をカウントし（ステップ1b05）、 i を1にリセットして（ステップ1b06）、 i 番目の文書に出現したタームの重みを式2に基づき計算し、ある閾値以上の重みを持つタームを i 番目の文書の重要タームとしてタームベクトルを生成し、タームベクトル格納部2cに格納する（ステップ1b07）。

【0070】さらに i を1増やして（ステップ1b0

8)、i が文書データ格納部 2 a に格納された文書の数よりも小さければ(ステップ 1 b 0 9)、ステップ 1 b 0 7に戻る。i が文書数を超えれば処理を終了する。

【0071】次にシソーラス概観生成処理の詳細な手順を説明する。図 19 は、図 16 (a) のシソーラスブラウジング用データ生成処理におけるシソーラス概観生成処理の詳細な処理手順例を示すフローチャートである。本図 19 に示すように、シソーラス概観生成処理は、まずシソーラス概観の要素となる代表タームを抽出し(ステップ 1 0 3 1)、抽出した代表タームをクラスタリングしてタームクラスタを生成し(ステップ 1 0 3 2)、生成したタームクラスタをシソーラス概観格納部 2 d に格納する(ステップ 1 0 3 3)。

【0072】以下、ステップ 1 0 3 1 の代表ターム抽出処理及びステップ 1 0 3 2 のタームクラスタ生成処理について詳細に説明する。シソーラス概観生成処理における代表ターム抽出処理の手順を、まずその概要について述べ、次にフローチャートを用いて詳細に説明する。

【0073】ある分野における一般的なタームを抽出する手がかりとして、当該分野におけるタームの出現頻度や出現文書数などのデータがある。これらをそのまま用いて、当該分野中、出現頻度の大きいターム、または出現文書数の大きいタームを一般的なタームとして抽出する方法が考えられるが、意味のないタームが抽出される

ターム重み計算式

$$w_{ik} = \frac{tf_{ik} \cdot \log(N/n_k)}{\sqrt{\sum_{j=1}^i (tf_{ij})^2 \cdot (\log(N/n_j))^2}}$$

w_{ik} : i 番目の文書における k 番目のタームの重み

tf_{ik} : i 番目の文書における k 番目のタームの出現頻度

N : 文書数

n_k : k 番目のタームが出現した文書数

尚、数 2 の式中分母は $tf \cdot idf$ の値をタームが出現した文書のサイズで正規化するための値である。

【0077】代表ターム抽出処理では、タームベクトルを構成するタームを抽出して、それぞれ重要タームとなった文書の数をカウントし、多くの文書で重要タームとなったタームを一般的なタームとして抽出する。次に、代表ターム抽出処理の詳細な処理手順例について図 20 を用いて説明する。

【0078】図 20 は、代表ターム抽出処理の詳細な処理手順例を示すフローチャートである。図 20 に示すように代表ターム抽出処理では、まず文書データ格納部 2 a に格納された文書の数を示す変数 i に 1 をセットして初期化し(ステップ 1 c 0 1)、i 番目の文書のタームベクトルを構成する重要タームをタームベクトル格納部 2 c から取得して、重要タームが既に代表タームリスト 3 c にあれば文書数 3 c 0 2 を 1 増やし、なければ文書数 3 c 0 2 を 1 として代表タームリスト 3 c に格納する

可能性がある。例えば、「経済」は経済分野において出現頻度は高いと予想されるが意味があるとは考えにくい。

【0074】これに対して本例の代表ターム抽出処理では、タームの重要度を利用することにより意味のある一般的なタームを抽出する。以下、具体的に説明する。本例の代表ターム抽出処理においては、予め当該分野に属する文書データベースから抽出した「タームベクトル」を利用する。この「タームベクトル」とは、文書の特徴付けるタームのリストであり、「Salton, G., et al.: A Vector Space Model for Automatic Indexing, Communications of the ACM, Vol.18, No.11(1975).」に記載の $tf \cdot idf$ 法 (Term Frequency inverse Document Frequency) を利用することにより抽出可能である。

【0075】この $tf \cdot idf$ 法は、文書のインデキシング方法として最もよく知られているものの一つであり、ある文書におけるタームの出現頻度 (tf) と、当該タームが出現した文書数の逆数 (idf) をかけた値を当該文書におけるタームの重要度とし、当該文書において重要度の高いターム (以下、重要ターム) を抽出してタームベクトルとする技術である。

【0076】文書の重要度を計算する式を数 2 に示す。

【数 2】

(ステップ 1 c 0 2)。

【0079】そして i を 1 増やして(ステップ 1 c 0 3)、i が文書数より小さければ(ステップ 1 c 0 4)ステップ 1 c 0 2に戻る。i が文書数を超えれば、代表タームリスト 3 c を文書数の降順にソートし(ステップ 1 c 0 5)、代表タームリスト 3 c の上位から指定された数だけ残して、後の重要タームを削除して処理を終了する(ステップ 1 c 0 6)。

【0080】次に、シソーラス概観生成処理におけるタームクラスタ生成処理の説明として、まずそのタームクラスタ生成処理の概要について述べ、次にフローチャートを用いて詳細に説明する。タームの集合からクラスタを生成するためにタームクラスタ生成処理では、まず、クラスタリング対象のタームを $\{t_1, t_2, \dots, t_m\}$ とし、初期状態として n 個のクラスタ $c_1 = \{t_1\}$ 、 $c_2 = \{t_2\}$ 、 \dots 、 $\{t_m\}$ を生成する。

【0081】そしてクラスタ総数が指定された数と等し

くなるまで関連度が最大であるタームクラスタの対を選択し、一つのタームクラスタにマージする処理を繰り返す。ここで二つのクラスタCとDの関連度 $rel(C, D)$ は、例えば、Cに属するターム t とDに属するターム s の関連度 $R(t, s)$ のうち、最大値をタームクラスタの関連度 $rel(C, D)$ とするなどの技術がある。

【0082】次に、タームクラスタ生成処理の詳細な処理手順について図21を用いて説明する。図21は、図16(a)のシソーラスブラウジング用データ生成処理におけるタームクラスタ生成処理の詳細な処理手順例を示すフローチャートである。

【0083】本図21に示すようにタームクラスタ生成処理では、まずターム間の関連度を格納した相関行列を生成する(ステップ1e01)。次に1つのタームからなるタームクラスタを生成して出力タームクラスタリスト3fを初期化し(ステップ1e02)、タームクラスタ数を示す変数Nにクラスタリングするターム数をセットして初期化する(ステップ1e03)。

【0084】N個のタームクラスタのうち関連度の高いタームクラスタの組C1及びC2を取得して(ステップ1e04)、マージし(ステップ1e05)、Nを1減らす(ステップ1e06)。Nが指定されたタームクラスタ数と同じでないなら(ステップ1e07)ステップ1e04に戻る。同じであれば処理を終了する。

【0085】ここで、ステップ1e01の相関行列生成処理と、ステップ1e02のタームクラスタ初期化処理、1e04のタームクラスタ選択処理、及びステップ1e05のタームクラスタマージ処理について、それぞれフローチャートを用いて詳細に説明する。まず、タームクラスタ生成処理における相関行列作成処理の詳細な説明を行なう。

【0086】図22は、図21のタームクラスタ生成処理における相関行列生成処理の詳細な処理手順例を示すフローチャートである。尚、以下の相関行列生成処理のフローチャートの説明において、タームリストとは、入力代表タームリスト3cのとき代表タームリスト、入力代表タームリスト3dのときは関連タームリストを指す。

【0087】本図22に示すように相関行列生成処理では、まず、相関行列の行番号を示す変数iに1をセットして初期化し(ステップ1e011)、相関行列のi行i列目の要素に0をセットする(ステップ1e012)。次に相関行列の列(又は行)番号を示すjにi+1をセットし(ステップ1e013)、タームリストのi番目のタームとj番目のタームの関連度を関連シソーラスから取得して(ステップ1e014)、相関行列i行j列目、及びj行i列目にその値をセットする(ステップ1e015)。

【0088】jを1増やして(ステップ1e016)、

jがタームリストに格納されたターム数を超えれば(ステップ1e017)、ステップ1e018に進み、超えなければステップ1e014に戻る。ステップ1e018ではiを1増やして、iがタームリストに格納されたターム数を超えれば(ステップ1e019)処理を終了し、超えなければステップ1e012に戻る。

【0089】次に、タームクラスタ生成処理におけるタームクラスタ初期化処理の詳細な説明を行なう。図23は、図21のタームクラスタ生成処理におけるタームクラスタ初期化処理の詳細な処理手順例を示すフローチャートである。本図23に示すようにタームクラスタ初期化処理ではまず、相関行列の行番号を示す変数iに1をセットして初期化し(ステップ1e021)、タームリストi番目のタームを出力タームクラスタリスト3fのi番目のタームリスト3f03にセットする(ステップ1e022)。

【0090】相関行列i行目に格納された関連度のうち最大値を出力タームクラスタリスト3fのi番目の最大値3f02にセットし(ステップ1e023)、当該最大値をとる列の番号を出力タームクラスタリスト3fのi番目のクラスタID3f01にセットする(ステップ1e024)。iを1増やして(ステップ1e025)、iがタームリストに格納されたターム数を超えれば(ステップ1e026)処理を終了し、超えなければステップ1e022に戻る。

【0091】この処理の具体的な結果が、図14に示す例である。尚、この図14に示す例は、図12(b)に示す関連タームリスト3dに基づくものである。すなわち、関連タームリスト3dにおける第1番目の行のターム「IMF」を出力タームクラスタリスト3fの1行目のタームリスト3f03にセットし、相関行列3eの1行目における最大の関連度である「14.4」を最大関連度3f02に、また、その最大関連度の列番号「3」をクラスタID3f03にセットする。同様に、出力タームクラスタリスト3fの2行目には「ドル」、「7.3」、「1」が、3行目には「ユーロ」、「14.4」、「1」がセットされる。

【0092】次に、タームクラスタ生成処理におけるタームクラスタ選択処理の詳細な説明を行なう。図24は、図21のタームクラスタ生成処理におけるタームクラスタ選択処理の詳細な処理手順例を示すフローチャートである。

【0093】図24に示すようにタームクラスタ選択処理ではまず、出力タームクラスタリスト3fのインデックスを示す変数iに1を、更に最も関連度の高い組として選択するタームクラスタのインデックスを示す変数C1とC2にそれぞれ0を、その関連度を格納する変数Maxに0をセットして初期化する(ステップ1e041)。そして出力タームクラスタリスト3fのi番目に格納された最大値とMaxの値を比較する(ステップ1

e042)。

【0094】Maxの方が大きければステップ1e044に進み、Maxの方が小さければ、出力タームクラスタリスト3fのi番目の最大値をMaxにセットし、またC1にiを、C2に出力タームクラスタリスト3fのi番目のクラスタIDをセットする(ステップ1e043)。

【0095】ステップ1e044ではiを1増やして、iが出力タームクラスタリスト3fに格納されたタームクラスタ数を超えれば(ステップ1e045)、処理を終了する。超えなければ、ステップ1e042に戻る。

【0096】次にタームクラスタ生成処理におけるタームクラスタマージ処理の詳細な説明を行なう。図25は、図21のタームクラスタ生成処理におけるタームクラスタマージ処理の詳細な処理手順例を示すフローチャートである。尚、以下のフローチャートの説明においてC1及びC2はマージするタームクラスタのインデックスである。

【0097】図25に示すようにタームクラスタマージ処理ではまず、相関行列3eの列番号を示す変数iに1をセットして初期化する(ステップ1e051)。次に相関行列C1行i列の関連度と相関行列C2行i列の関連度を比較し(ステップ1e052)、相関行列C2行i列の関連度の方が小さければステップ1e054に進み、大きければ、その値を相関行列C1行i列の要素にセットする(ステップ1e053)。

【0098】例えば、図13の相関行列3eの4列目における1行目(C1)と3行目(C2)では、それぞれの関連度は「4.1」と「8.6」であり、相関行列3eの4列目における1行目(C1)において、「4.1」を削除して「8.6」にセットし、3行目(C2)の「8.6」を「0」にセットする。

【0099】ステップ1e054では、相関行列C2行i列目の要素に0をセットし、iを1増やして(ステップ1e055)、iがタームリストのターム数を超えれば(ステップ1e056)、ステップ1e057に進み、超えなければ、ステップ1e052に戻る。

【0100】ステップ1e057では、出力タームクラスタリスト3fのC1番目のタームリストに、出力タームクラスタリストC2番目のタームリストに格納されたタームを追加し、さらに相関行列3eのC1行目に格納された関連度のうちの最大値となる列の番号を、出力タームクラスタリスト3fのC1番目の「最大関連度のクラスタID3f01」に、また、その値を「最大関連度3f02」に格納する。そして出力タームクラスタリスト3fのC2番目の「タームリスト3f03」の情報を削除し、「最大値をとるタームクラスタID3f01」と「最大関連度3f02」に「0」をセットして(ステップ1e058)処理を終了する。

【0101】例えば、図14の出力タームクラスタリス

ト3fにおいて、1番目(1行目)のタームリスト3f03(「IMF」)に2行目のタームリストの「ドル」を追加し、最大関連度のクラスタID3f01に「3」を、また最大関連度3f02に「14.4」を格納し、2行目のタームリストの「ドル」を削除し、最大関連度のクラスタID3f01および最大関連度3f02に「0」を格納する。この操作を繰り返し、また、タームリストが削除されて最大関連度のクラスタID3f01および最大関連度3f02に「0」が格納された行を削除することにより、図14(b)に示す内容となる。尚、各行におけるターム数は「10」等、所定の数とする。

【0102】次に、ソーラスブラウジング処理における関連ターム取得処理に関して、まずその概要について述べ、次にフローチャートを用いて詳細を説明する。ユーザが入力タームクラスタリスト3gから選択したタームクラスタに属するターム(以下、「種ターム」と呼ぶ)の関連タームは、関連ソーラスを検索することによって取得できる。

【0103】取得された関連タームの総数が小さければ(例えば300個以内)、それらすべてをクラスタリングの対象すれば良い。一方で、その数が多いときは、一例として各種タームから均等に関連ターム取得する。すなわち、すべての種タームから関連タームを関連度の強い順にXタームずつ取得する。これにより、クラスタリングするタームの数が予め定めた数の範囲内となるので、タームクラスタの生成のための計算量を抑えることが可能となる。

【0104】関連ターム取得処理の詳細な処理手順について図26を用いて説明する。図26は、図16(b)のソーラスブラウジング処理における関連ターム取得処理の詳細な処理手順例を示すフローチャートである。本図26に示すように関連ターム取得処理では、まず、入力タームクラスタリスト3gのタームリスト3g01に格納されたタームを関連タームリスト3dに格納する。ここで、各タームを格納した行の順位3d02は1とする(ステップ1d01)。例えば、15行目の「東京外為市場」、20行目の「ドル」などである。

【0105】そして、入力タームクラスタリスト3gのタームリスト3g01に格納されたタームの総数を表す変数iに1をセットして初期化し(ステップ1d02)、入力タームクラスタリスト3gのタームリスト3g01に格納されたタームのうち、i番目のタームの関連タームを関連ソーラス格納部2bを検索して取得し、それらを関連度の降順にソートして順位付けして関連タームリスト3dの末尾に格納する(ステップ1d03)。その結果、例えば、図12(a)の関連タームリスト3dにおける15~18行目のようになる。

【0106】さらにiを1増やして(ステップ1d04)。iがタームリスト3g01に格納されたタームの

総数を超えれば（ステップ1d05）、ステップ1d06に進み、超えなければステップ1d03に戻る。ステップ1d06では関連タームリスト3dを順位の昇順でソートする。この結果、図12（b）における関連タームリスト3dのようになるが、この時点では、重複した関連タームが存在している可能性があるので、さらに、以下の処理を行なう。

【0107】すなわち、関連タームリスト3dのインデックスを示す変数iに1を、関連タームのうちクラスタリング対象として選択したタームの数を表す変数jに0をセットして初期化する（ステップ1d07）。そして、関連タームリスト3dのi番目のタームがリストの1～i-1番目に無ければ（ステップ1d08）、jを1増やして（ステップ1d09）ステップ1d11に進む。関連タームリスト3dにあれば、関連タームリスト3dのi番目の順位を0にセットする（ステップ1d10）。ステップ1d11ではiを1増やして（ステップ1d11）、iが関連タームリスト3dに格納された関連ターム数を超えれば（ステップ1d12）、ステップ1d15へ進む。

【0108】それ以外は、関連タームリスト3dのi-1番目の順位とi番目の順位が同じであれば（ステップ1d13）、ステップ1d08に戻る。同じでなければステップ1d14に進み、jがクラスタリングするタームの数と同じになれば、i番目以降の関連ターム及び順位が0の関連タームを削除して（ステップ1d15）、処理を終了する。しかし、jがクラスタリングするタームの数と同じでないなら、ステップ1d08に戻る。このようにして、図12（b）の関連タームリスト3dは重複の無いものとなる。

【0109】以上、本例のシソーラスブラウジング方法の処理手順について説明した。以下、このようなシソーラスブラウジングに係わる具体的な画面に関して説明する。図27は、図1におけるシソーラスブラウジングシステムで表示される画面の構成例を示す説明図である。本画面は、図6に示す経済分野の関連シソーラスをブラウジングしている例であり、図27（a）はシソーラス概観の表示状態、図27（b）はタームクラスタのズーム状態である。

【0110】図27（a）に示すように経済分野のシソーラス概観として、「景気」、「売り上げ」、「消費者」などからなるタームクラスタや、「税制」、「所得」、「国税庁」等からなるタームクラスタなどが表示されている。ここでユーザが第3番目のタームクラスタに興味を持ち、当該タームクラスタを選択し、<ズーム>ボタンをクリックしてズームを指示すると、画面は図27（b）に示す状態となる。

【0111】図27（b）は、図27（a）中の第3のタームクラスタを選択してズームした状態であり、「IMF」、「ドル」、「外貨」等からなるタームクラ

スタや、「円」、「東京外為市場」、「円買い」等からなるタームクラスタ等が表示されている。このようにユーザはある特定のタームを入力せずとも、関連シソーラスを効率よくブラウジングすることができる。

【0112】次に、このような実施形態の第1～3の変形例を説明する。まず、第1の変形例として、代表ターム抽出処理の変形例を説明する。第1の変形例にかかるシソーラスブラウジングシステムのハードウェア構成は図4に示した実施形態と同じである。またモジュール構成は、図1におけるモジュール構成からタームベクトル抽出部1b及びタームベクトル格納部2bを除いた形態と同じである。

【0113】第1の変形例において代表ターム抽出部1cの変形例は、文書の構成要素を手がかりにして代表タームを抽出する。具体的には、文書の特徴付けるタームが出現しやすい構成要素からタームを抽出して、その出現文書数が多いタームを代表タームとして代表タームリスト3cに格納する。例えば、文書データ格納部2aに格納されているデータが新聞記事であれば、第1文すなわち見出しに出現したタームを抽出し、また、特許明細書であれば、要約書又は従来の技術と課題からタームを抽出して重要タームとする。

【0114】この第1の変形例における代表ターム抽出部1cの変形例によれば、一般的なタームを含む文書の構成要素が明らかな新聞記事のような場合に、効率よくシソーラス概観を生成することができる。

【0115】次に、第2の変形例として、ユーザによるタームクラスタの編集処理を含むシソーラスブラウジングシステムについて説明する。第2の変形例にかかるシソーラスブラウジングシステムのハードウェア構成及びモジュール構成は図4及び図1に示すものと同じである。第2の変形例において、関連ターム取得部1dに入力される入力タームクラスタ3gは、表示されたタームクラスタのうち、ユーザが選択してタームを追加・削除したものである。

【0116】例えば、図27に示す画面例において、「ズーム」ボタンの隣に「編集」ボタンを設け、ユーザが第3番目のタームクラスタ（「円、為替、日銀、外貨」、・・・）と、「編集」ボタンを選択すると、第3番目のタームクラスタの各ターム（「円」、「為替」、「日銀」、「外貨」、・・・）の一覧表と、「追加」ボタン、「削除」ボタンが設けられた「編集画面」を別ウィンドウに表示する。

【0117】ユーザは、この「編集画面」上で「追加」ボタンと「削除」ボタンを操作して、一覧表の各タームの内、必要なものだけ、例えば、「円」、「為替」、「日銀」のみを残した新たなタームクラスタを作成する。そして、ユーザが「終了」ボタン等を選択すると、新たなタームクラスタに対するズームを行なう。

【0118】このように、第2の変形例によれば、入力

タームクラスタ3gに対するユーザの編集処理が可能となり、ユーザはより自分の興味に合わせてシソーラスをズームングすることができ、使い勝手の良いシソーラスのブラウジングを達成することができる。

【0119】さらに、第3の変形例として、ユーザが選択したタームクラスタのズームング結果を表示する際、種ターム（選択した元のタームクラスタに属するターム）は他の関連タームとは区別可能な状態で表示するシソーラスブラウジング技術について説明する。

【0120】第3の変形例にかかるシソーラスブラウジングシステムのハードウェア構成及び処理手順の概要は、図1～図27で説明したものと同様である。本第3の変形例において、生成された出力タームクラスタ3fは、種タームと他のタームを区別可能なよう、色、もしくはフォント等を変えて表示する。ここで種タームとは、タームクラスタリングの入力となったタームクラスタに属するタームであり、シソーラス概観から現在表示中のタームクラスタに至るまでにユーザが選択したタームクラスタのいずれかに属する。

【0121】このようにするとにより、ブラウジングの履歴を階層的に表示可能な階層シソーラスとは異なり、“迷子”になりやすいネットワーク型のシソーラスにおいても、ユーザは現在の状態を参照するだけでこれまでのどのようなタームを辿ったかを容易に知ることができる。

【0122】以上、ある分野に属する文書データから自動的に生成した関連シソーラスをブラウジングするシソーラスブラウジングシステムと方法について説明した。次に本発明の第2の実施形態例として、ユーザによるキータームの入力がある場合のシソーラスブラウジング技術について説明する。

【0123】まず、図28を用いて本第2の実施形態にかかるシソーラスブラウジングシステムのモジュール構成を示し、次に図29を用いてシソーラスブラウジング方法の処理手順について述べ、最後に図30を用いて本第2の実施形態におけるシソーラスブラウジングの画面の一例を示す。尚、本第2の実施形態にかかるシソーラスブラウジングシステムのハードウェア構成は図4に示す第1の実施形態と同じで良い。

【0124】図28は、本発明のシソーラスブラウジングシステムの本発明に係る構成の第2の実施形態例を示すブロック図である。図28に示すように、本シソーラスブラウジングシステムは、処理部として関連ターム取得部1d及びタームクラスタ生成部1e、データ格納部として関連シソーラス格納部2b、データとしてターム3h、関連タームリスト3d、相関行列3e、及び出力タームクラスタリスト3fから構成される。

【0125】これらのうち、ターム3hはユーザが入力したタームである。それ以外の構成要素の概要は図1に示す例と同じである。このような構成のシソーラスブラ

ウジングシステムの処理動作を図29と図30を用いて説明する。

【0126】図29は、図28におけるシソーラスブラウジングシステムの本発明に係る処理手順例を示すフローチャートである。図29に示すように、まず関連シソーラスを検索して、ユーザが入力したターム3hの関連タームを取得し関連タームリストに格納し（ステップ201）、関連タームリストに格納されたタームをクラスタリングして（ステップ202）、タームクラスタをユーザに表示する（ステップ203）。

【0127】これらの処理のうち、ステップ201の関連ターム取得処理は、1つのタームからなるタームクラスタがユーザから選択されたとすれば、第1の実施形態例における関連ターム取得処理と同様であるので、説明を省略する。同様に、ステップ202におけるタームクラスタ生成処理は、第1の実施形態例におけるタームクラスタ生成処理と同じであるので、説明を省略する。

【0128】図30は、図28におけるシソーラスブラウジングシステムで表示される画面の構成例を示す説明図である。図30で示す画面は、図6に示す経済分野の関連シソーラスにキーターム「円高」を入力してブラウジングしている例である。本図30に示すように「円高」の関連タームには「ドル安」、「東京外為市場」等があるが、それらは関連の強いものがまとまって表示されている。

【0129】例えば、第1のタームクラスタとして「ドル安」、「貿易不均衡」、「黒字減らし」等、第2のタームクラスタとして「東京外為市場」、「円買い」、「差益」があることがわかる。このように関連タームがいくつかのタームクラスタに分割されて表示されるので、ユーザは効率よく関連タームを参照することができる。以上、第2の実施形態例として、ユーザによるタームの入力がある場合の関連シソーラスのブラウジングについて説明した。

【0130】以上、図1～図30を用いて説明したように、本実施例のシソーラスブラウジングシステムと方法では、一般的なタームから構成されるシソーラスの概観（全体構造）を表示し、このシソーラスの概観からより限定的なタームを含む部分構造へユーザをナビゲートする。これにより、ユーザは、関連シソーラスの概観を容易に把握でき、更に興味を持ったタームクラスタのズームングにより、効率的に関連シソーラスをブラウジングすることができる。

【0131】すなわち、関連シソーラスの検索要求が漠然としている、或いは検索対象の関連シソーラスについて熟知していない等の理由で適切な検索タームを入力できないユーザでも、表示されたシソーラスの概観を参照すれば全体の構造を把握でき、更に興味を持ったタームクラスタを選択してズームングを指示すれば、シソーラスをブラウジングすることができる。これにより、全体

の構造を把握しにくく、かつブラウジング中に迷子になりやすいネットワーク型の関連シソーラスを効率的に参照することができる。

【0132】また、関連シソーラスを文書データベースから生成した場合、従来の文書データベース検索では不可能であったデータベースそのもののブラウジングが可能となる。すなわち、従来、文書データベースを参照する技術として、格納されている文書を検索して本文を参照する、格納されているタイトル等の書誌情報の一覧を参照する等があるが、これらの技術はいずれも格納されている文書それぞれに対応するデータを参照するものであり、文書データベースに格納された個々のデータよりも全体の特徴を把握したいユーザ（例えば、ある年度の特許データベースの傾向を知りたいユーザ）にとって有用な技術とは言えなかった。

【0133】これに対して、本例のように、文書データベースから自動的に生成した関連シソーラスは当該データベースに特徴的な関連関係を抽出可能であることから、関連シソーラスをブラウジングすれば当該データベースの特徴を把握することが可能となる。例えば、ある年度の特許データベースから生成した関連シソーラスは「音声認識」と「カーナビゲーション」の関連度が強いなどを把握できる。また、従来技術は関連シソーラスの部分構造のみを表示するものであったが、本例では、関連シソーラス全体をブラウジングすることができる。

【0134】尚、本発明は、図1～図30を用いて説明した実施例に限定されるものではなく、その要旨を逸脱しない範囲において種々変更可能である。例えば、本例では、経済分野の文書データの関連シソーラスについて説明したが、技術文書のデータベースや特許データベースなどの関連シソーラスについても同様に適用することができる。

【0135】

【発明の効果】本発明によれば、関連シソーラス中の一般的なタームで構成されたシソーラスの概観をユーザに提供し、更にユーザが興味を持った部分構造をより詳細に表示することができ、従来の文書データベース検索とは異なる「文書データベースのブラウジング」という新しい情報処理技術をユーザに提供することが可能となり、ユーザは表示されたシソーラスの概観から当該文書データベースの全体的な特徴をつかみ、更に興味を持った部分をズームングすることにより当該文書データベースの詳細な特徴について知ることができ、関連シソーラスの利用効率を向上させることが可能である。

【図面の簡単な説明】

【図1】本発明のシソーラスブラウジングシステムの本発明に係る構成の第1の実施形態例を示すブロック図である。

【図2】関連シソーラスの一例を示す説明図である。

【図3】図1におけるシソーラスブラウジングシステム

の本発明に係る処理の概要を示す説明図である。

【図4】図1におけるシソーラスブラウジングシステムのハードウェア構成例を示すブロック図である。

【図5】図1における文書データ格納部の構成例を示す説明図である。

【図6】図1における関連シソーラス格納部の構成例を示す説明図である。

【図7】図1におけるタームベクトル格納部の構成例を示す説明図である。

【図8】図1におけるシソーラス概観格納部の構成例を示す説明図である。

【図9】図1における共起タームテーブルの構成例を示す説明図である。

【図10】図1における文書タームテーブルの構成例を示す説明図である。

【図11】図1における代表タームリストの構成例を示す説明図である。

【図12】図1における関連タームリストの構成例を示す説明図である。

【図13】図1における相関行列の構成例を示す説明図である。

【図14】図1における出力タームクラスタリストの構成例を示す説明図である。

【図15】図1における入力タームクラスタリストの構成例を示す説明図である。

【図16】本発明のシソーラスブラウジング方法の処理手順例を示すフローチャートである。

【図17】図16（a）のシソーラスブラウジング用データ生成処理における関連シソーラス生成処理の詳細な処理手順例を示すフローチャートである。

【図18】図16（a）のシソーラスブラウジング用データ生成処理におけるタームベクトル抽出処理の詳細な処理手順例を示すフローチャートである。

【図19】図16（a）のシソーラスブラウジング用データ生成処理におけるシソーラス概観生成処理の詳細な処理手順例を示すフローチャートである。

【図20】代表ターム抽出処理の詳細な処理手順例を示すフローチャートである。

【図21】図16（a）のシソーラスブラウジング用データ生成処理におけるタームクラスタ生成処理の詳細な処理手順例を示すフローチャートである。

【図22】図21のタームクラスタ生成処理における相関行列生成処理の詳細な処理手順例を示すフローチャートである。

【図23】図21のタームクラスタ生成処理におけるタームクラスタ初期化処理の詳細な処理手順例を示すフローチャートである。

【図24】図21のタームクラスタ生成処理におけるタームクラスタ選択処理の詳細な処理手順例を示すフローチャートである。

【図 25】図 21 のタームクラスタ生成処理におけるタームクラスタマージ処理の詳細な処理手順例を示すフローチャートである。

【図 26】図 16 (b) のソーラスブラウジング処理における関連ターム取得処理の詳細な処理手順例を示すフローチャートである。

【図 27】図 1 におけるソーラスブラウジングシステムで表示される画面の構成例を示す説明図である。

【図 28】本発明のソーラスブラウジングシステムの本発明に係る構成の第 2 の実施形態例を示すブロック図である。

【図 29】図 28 におけるソーラスブラウジングシステムの本発明に係る処理手順例を示すフローチャートである。

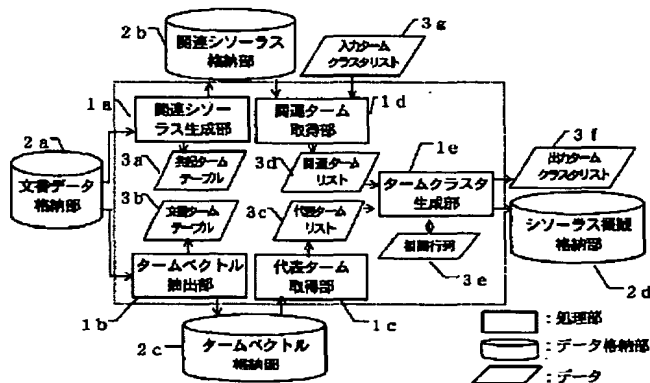
【図 30】図 28 におけるソーラスブラウジングシステムで表示される画面の構成例を示す説明図である。

【符号の説明】

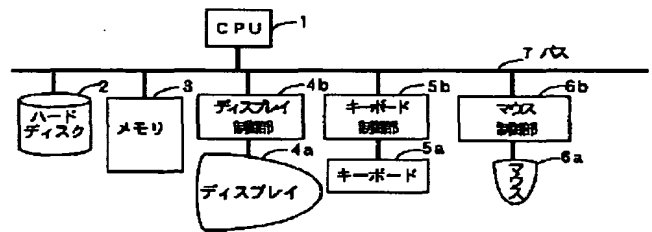
1 : CPU、2 : ハードディスク、3 : メモリ、4 a : ディスプレイ、4 b : ディスプレイ制御部、5 a : キーボード、5 b : キーボード制御部、6 a : マウス、6

b : マウス制御部、7 : バス、1 a : 関連ソーラス生成部、1 b : タームベクトル抽出部、1 c : 代表ターム取得部、1 d : 関連ターム取得部、1 e : タームクラスタ生成部、2 a : 文書データ格納部、2 b : 関連ソーラス格納部、2 c : タームベクトル格納部、2 d : ソーラス概観格納部、3 a : 共起タームテーブル、3 b : 文書タームテーブル、3 c : 代表タームリスト、3 d : 関連タームリスト、3 e : 相関行列、3 f : 出力タームクラスタリスト、3 g : 入力タームクラスタリスト、3 h : ターム、2 a 01 : 文書データ、2 b 01 : ターム X、2 b 02 : ターム Y、2 b 03 : 関連度、2 c 01 : 文書 ID、2 c 02 : 重要タームリスト、2 d 01 : タームリスト、3 a 01 : ターム X、3 a 02 : ターム Y、3 a 03 : 共起頻度、3 b 01 : 文書 ID、3 b 02 : ターム、3 b 03 : 出現頻度、3 c 01 : 代表ターム、3 c 02 : 文書数、3 d 01 : 関連ターム、3 d 02 : 順位、3 f 01 : 最大関連度のクラスタ ID、3 f 02 : 最大関連度、3 f 03 : タームリスト、3 g 01 : タームリスト。

【図 1】



【図 4】



【図 5】

文書データ格納部の構成

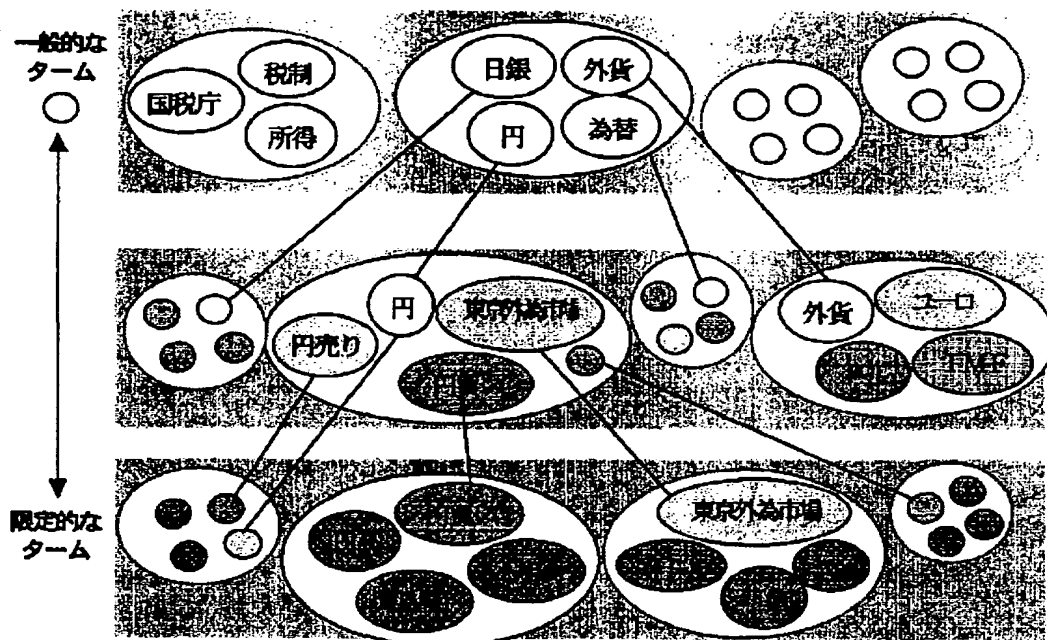
2 a 01	
#	文書データ
1	21日の東京外為市場の円相場は、下値の...
2	通貨不安解消へ日本政府に協力を要請...
	...
8	経済企画庁による景気動向見通しによれば...
9	IMF、週明けにもインドと融資交渉を...
10	...

【図 2】



【図3】

関連シソーラスのナビゲートの一例



【図6】

関連シソーラス格納部の構成

#	タームX	タームY	関連度
1	'円'	'東京外為市場'	11.5
2	'円'	'円買い'	10.7
3	'円'	'終値'	9.3
4	'円'	'最安値'	8.2
5		...	

【図7】

タームベクトル格納部の構成

#	文書ID	重要タームリスト
1	1	'円高' '東京外為市場' 'ドル安' ...
2	2	'アジア' '外貨' '円' '為替' ...
...
8	3	'所得' '円高' '税制' '景気' ...
9	9	'インド' '融資' 'IMF' ...
10

【図11】

代表タームリストの構成
(a)ソート前

#	代表ターム	文書数
1	'為替'	451
2	'所得'	172
3	'ロンドン市場'	63
4	'日銀'	338
5	'円'	402
6	'住宅ローン控除'	31
7	'外貨'	329
8	'国税庁'	197
...

【図8】

シソーラス概念格納部の構成

#	タームリスト
1	'景気' '売り上げ' '消費者' ...
2	'税制' '所得' '国税庁' ...
3	'円' '為替' '日銀' '外貨' ...
...	...
9	'債' '投資' '東証' ...
10	'失業' 'リストラ' '倒産' ...

【図9】

共起タームテーブルの構成

#	タームX	タームY	共起頻度
1	'円相場'	'東京外為市場'	259
2	'下値'	'東京外為市場'	67
3	'円相場'	'下値'	73
4	'円相場'	'円高'	142
5		...	

(b)ソート後

#	代表ターム	文書数
1	'株'	498
2	'為替'	451
3	'円'	402
4	'ドル'	386
5	'市場'	373
6	'投資'	361
7	'日銀'	338
8	'貿易'	334
...

【図10】

文書タームテーブルの構成

3b01	3b02	3b03	
書文ID	ターム	出現頻度	
1	1	「東京外為市場」	2
2	1	「円相場」	3
3	1	「下値」	1
4	1	「円高」	5
		...	
32	2	「通貨不安解消」	1

【図12】

関連タームリストの構成

(a) ソート前

3d01	3d02
関連ターム	順位
...	...
15	'東京外為市場' 1
16	'円高' 2
17	'株値' 3
18	'暴安値' 4
...	...
20	'ドル' 1
21	'ユーロ' 2
...	...

(b) ソート後

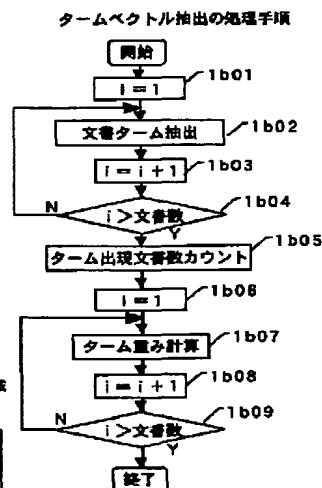
3d01	3d02
関連ターム	順位
1	'IMF' 1
2	'ドル' 1
3	'ユーロ' 1
4	'円' 1
...	...
9	'東京外為市場' 1
10	'外貨' 1
11	'円高' 2
...	...

【図13】

相関行列の構成

1	2	3	4	5	6	...	n
1	0.0	15.8	14.4	4.1	7.8	1.2	3.3
2	15.8	0.0	5.9	2.1	0.0	0.1	1.2
3	14.4	5.9	0.0	8.8	0.0	7.5	0.0
4	4.1	2.1	8.8	0.0	5.8	6.0	7.1
5	7.8	0.0	0.0	5.8	0.0	4.7	8.5
6	1.2	0.1	7.5	6.0	4.7	0.0	1.7
...
n	3.3	1.2	0.0	7.1	8.5	1.7	0.0

【図18】



【図14】

出力タームクラスタリストの構成

(a) 初期状態

	3f01	3f02	3f03
	最大関連度のタームID	最大関連度	タームリスト
1	3	14.4	'IMF'
2	1	15.8	'ドル'
3	1	14.4	'ユーロ'

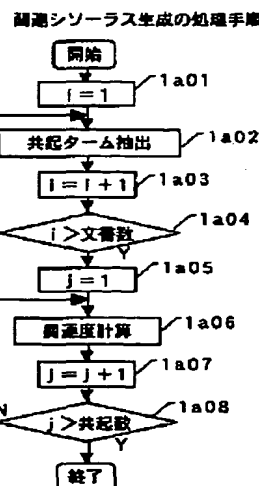
(b) クラスタ生成後の状態

【図15】

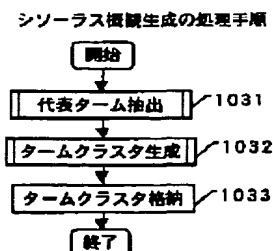
入力タームクラスタリストの構成

	3g01
2	タームリスト
1	'円' '為替' '日経' '外貨' ...
2	
3	
	...

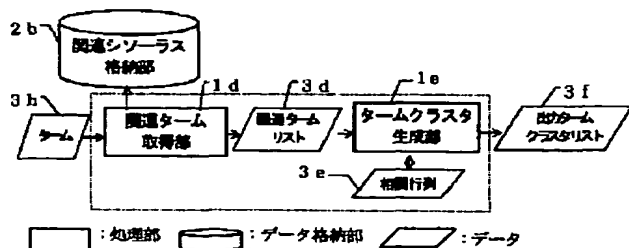
【図17】



【図19】

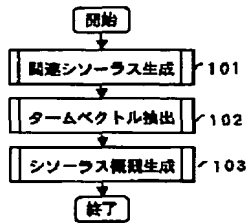


【図28】

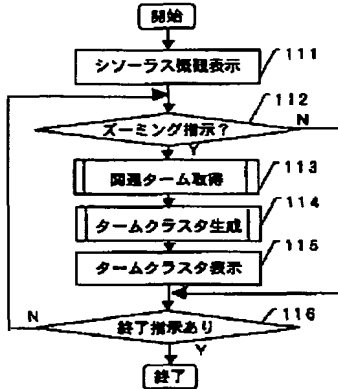


【図16】

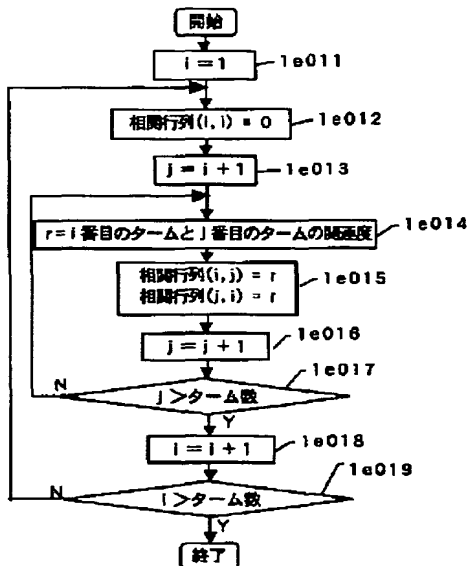
(a) シソーラスブラウジング用データ生成処理手順



(b) シソーラスブラウジング処理手順

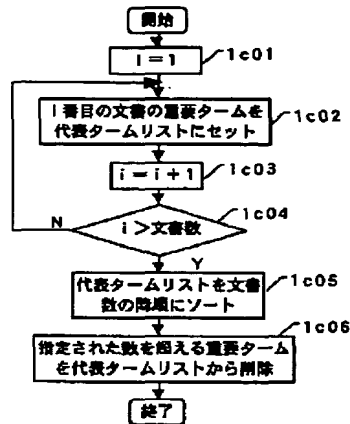


【図22】

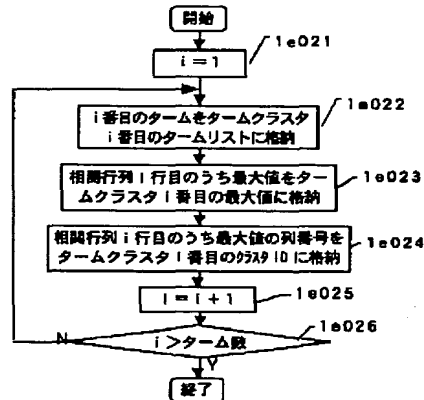
タームクラスタ生成部における
相関行列生成処理手順

【図20】

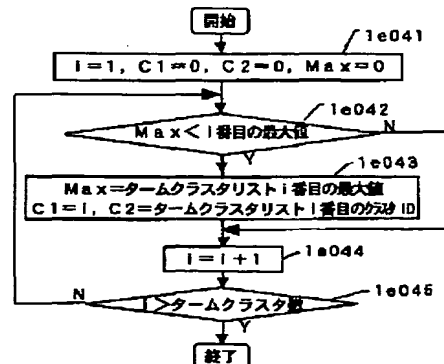
代表ターム抽出の処理手順



【図23】

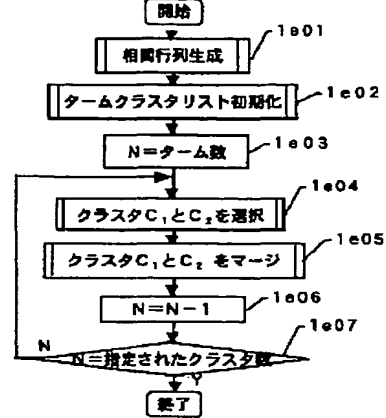
タームクラスタ生成部における
タームクラスタ初期化処理手順

【図24】

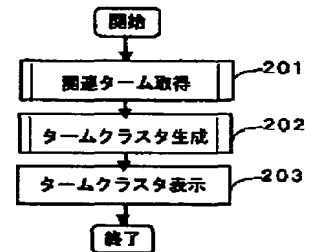
タームクラスタ生成部における
タームクラスタ選択処理手順

【図21】

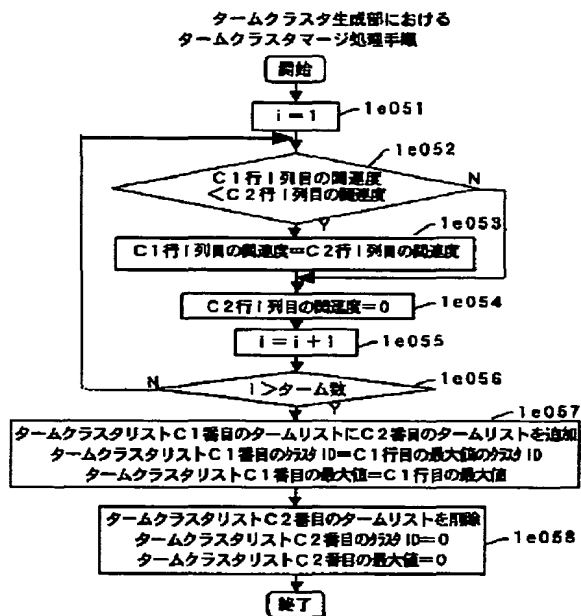
タームクラスタ生成処理手順



【図29】

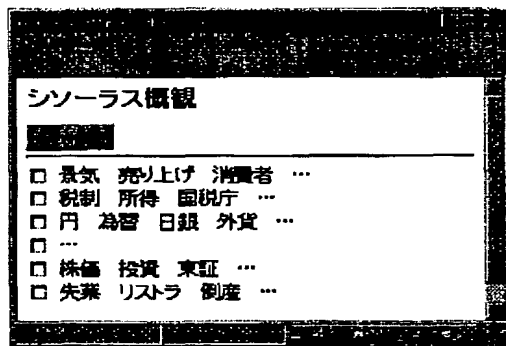


【図25】

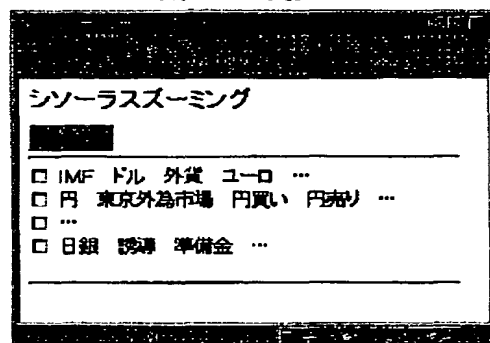


【図27】

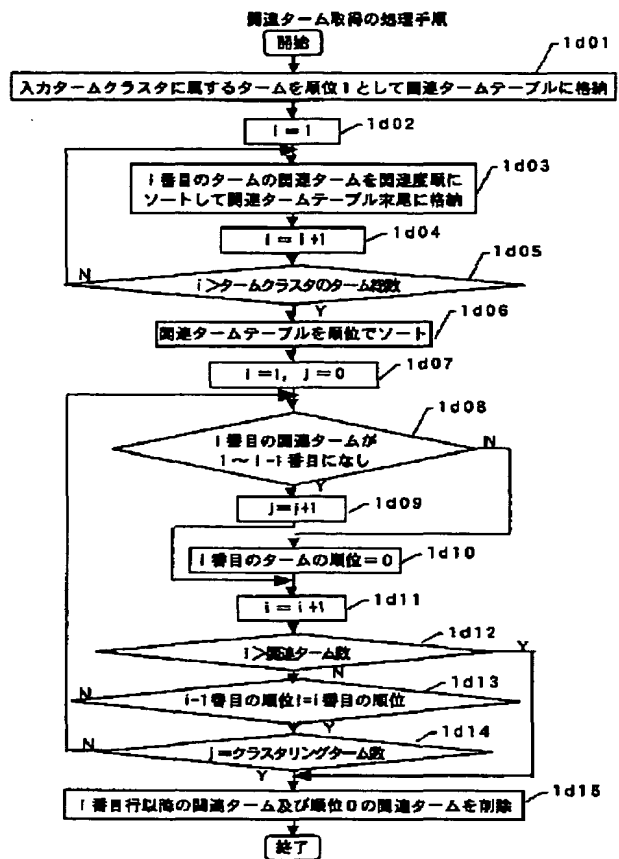
(a) シソーラス概観の表示画面例



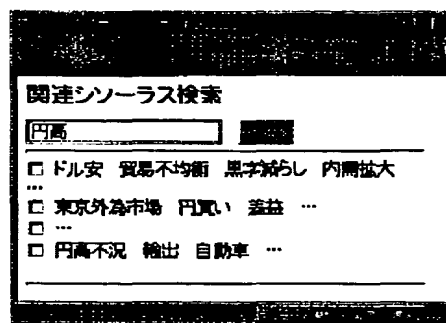
(b) ズーミング後



【図26】



【図30】



フロントページの続き

(72)発明者 森本 康嗣

東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72)発明者 山崎 紀之

神奈川県横浜市戸塚区戸塚町5030番地 株
式会社日立製作所ソフトウェア事業部内

(72)発明者 飯田 恵子

神奈川県横浜市戸塚区戸塚町5030番地 株
式会社日立製作所ソフトウェア事業部内

(72)発明者 内田 安彦

神奈川県横浜市戸塚区戸塚町5030番地 株
式会社日立製作所ソフトウェア事業部内

Fターム(参考) 5B075 ND02 ND35 NR02 NR12 PP13

PQ02 PQ42 PQ46 QP03

5E501 AC33 AC34 BA05 CA02 CB09

EA10 EB05 EB20 FA03 FA06